Proportional Size Distribution (PSD)

Alex J. Benecke

December 19, 2017

I will use a Chi-Square test to see if there is a difference between PSD between years (2013 - 2016).

Data Preparation

Load Data

```r
lmbs <- read.csv("Data/Clean-Data/2012-2016_nearshore-survey-largemouth-bass_Stock_CLEAN.csv") %>%
    filter(Year >= 2013) %>% arrange(Year, FID, Length)
lmbs$fyr <- as.factor(lmbs$fyr)

str(lmbs)
```

```
## 'data.frame':    412 obs. of  16 variables:
##  $ Year   : int  2013 2013 2013 2013 2013 2013 2013 2013 2013 2013 ...
##  $ Site   : int  11 11 11 11 11 11 11 10 10 10 ...
##  $ FID    : int  1 2 3 4 6 7 8 9 10 17 ...
##  $ Weight: num  NA NA NA NA NA NA NA NA NA NA ...
##  $ Length: int  395 348 266 224 318 273 426 387 264 291 ...
##  $ AC     : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ AGE    : int  3 2 1 1 2 1 3 4 1 1 ...
##  $ SexCon: int  3 8 8 6 8 8 3 8 3 8 ...
##  $ Sex    : int  1 2 2 2 2 2 1 2 1 2 ...
##  $ Delts  : logi  NA NA NA NA NA NA ...
##  $ logW   : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ logL   : num  2.6 2.54 2.42 2.35 2.5 ...
##  $ fyr    : Factor w/ 4 levels "2013","2014",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ Ws     : num  935 617 256 146 460 ...
##  $ Wr     : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ gcat   : Factor w/ 3 levels "preferred","quality",..: 1 2 3 3 2 3 1 1 3 3 ...
```

```r
headtail(lmbs)
```

```
##     Year  Site FID Weight Length AC AGE SexCon Sex Delts     logW     logL
## 1   2013    11   1     NA    395 NA   3      3   1    NA       NA 2.596597
## 2   2013    11   2     NA    348 NA   2      8   2    NA       NA 2.541579
## 3   2013    11   3     NA    266 NA   1      8   2    NA       NA 2.424882
## 410 2016    15 130    305    266  3   2      8   2    NA 2.484300 2.424882
## 411 2016    15 131    282    261  3   2      3   1    NA 2.450249 2.416641
## 412 2016 15972 132    971    395  3   7      3   1    NA 2.987219 2.596597
##      fyr       Ws       Wr      gcat
## 1   2013 934.6786       NA preferred
## 2   2013 617.4316       NA   quality
## 3   2013 256.2345       NA     stock
## 410 2016 256.2345 119.0316     stock
## 411 2016 240.8044 117.1075     stock
## 412 2016 934.6786 103.8860 preferred
```

```r
unique(lmbs$Year)  ### See that there is no 2012
```

```
## [1] 2013 2014 2015 2016
```

View Data

```r
(lmbs.LF <- xtabs(~Year+gcat,data=lmbs))
```

```
##       gcat
## Year   preferred quality stock
##   2013        16      41    41
##   2014        18      57    65
##   2015        15      38    14
##   2016        11      44    52
```

Chi-Squares Test

Is there a difference in the number of fish in each gabelhouse categorie during the years 2013 - 2016?

```r
chisq.test(lmbs.LF)
```

```
##
##  Pearson's Chi-squared test
##
## data:  lmbs.LF
## X-squared = 16.694, df = 6, p-value = 0.01048
```

This seems to suggest that the proportional stock distribution (PSD) is different for largemouth bass between years ($X^2 = 16.694$, df = 6, P = 0.01048).

In which years is PSD different?

```r
round(prop.table(lmbs.LF,margin=1)*100,0)
```

```
##       gcat
## Year   preferred quality stock
##   2013        16      42    42
##   2014        13      41    46
##   2015        22      57    21
##   2016        10      41    49
```

Remarkably the percent of quality fish is the same for 2014 and 2016 and a bit higher for 2015. the percentage of fish in each gcat is similar between years 2013, 2014, and 2016, However, the year 2015 appears to have a higher percentage of large fish and far fewer small fish.

1) Could this be some sort of sampling bias?
2) Could this be a result of sampling different sites? where the 2015 sites more suitable for LMB?
3) Are the years really different or do I just have too few to say for sure?

Compare PSD-Q between years 2013 - 2016

```r
lmbs %<>% mutate(gcatQ=mapvalues(gcat,
                                 from=c("stock","quality","preferred"),
                                 to=c("quality-", "quality+", "quality+")),
                 gcatQ=droplevels(gcatQ))

(lmb.LFQ <- xtabs(~Year+gcatQ,data = lmbs))
```

```
##       gcatQ
## Year   quality+ quality-
##   2013       57       41
##   2014       75       65
##   2015       53       14
##   2016       55       52
```

```r
chisq.test(lmb.LFQ)
```

```
##
##  Pearson's Chi-squared test
##
## data:  lmb.LFQ
## X-squared = 15.306, df = 3, p-value = 0.001573
```

```r
(ps.Q <- c(chisq.test(lmb.LFQ[1:2,])$p.value,    ### 2013-2014
           chisq.test(lmb.LFQ[c(1,3),])$p.value, ### 2013-2015
           chisq.test(lmb.LFQ[c(1,4),])$p.value, ### 2013-2016
           chisq.test(lmb.LFQ[2:3,])$p.value,    ### 2014-2015
           chisq.test(lmb.LFQ[c(2,4),])$p.value, ### 2014-2016
           chisq.test(lmb.LFQ[3:4,])$p.value))   ### 2015-2016

(p.val.Q <- p.adjust(ps.Q))
```

```
## [1] "13-14" "13-15" "13-16" "14-15" "14-16" "15-16"
```

```
##    Year p-value Adjusted p
## 1 13-14  0.5694     1.0000
## 2 13-15  0.0084     0.0337
## 3 13-16  0.4060     1.0000
## 4 14-15  0.0007     0.0036
## 5 14-16  0.8338     1.0000
## 6 15-16  0.0005     0.0027
```

The PSD-Q of largemouth bass is different for at least one of the years during 2013 - 2016 (Chi-Squared, $X^2 = 15.3$, p = 0.0016). The adjusted p-values show a difference in PSD-Q between years 2013 - 2015 (p = 0.0337), 2014 - 2015 (p = 0.0036), and 2015 - 2016 (p = 0.0027).

Compare PSD-P between years 2013 - 2016

```r
lmbs %<>% mutate(gcatP=mapvalues(gcat,
                                 from=c("stock","quality","preferred"),
                                 to=c("preferred-", "preferred-", "preferred+")),
                 gcatP=droplevels(gcatP))

(lmb.LFP <- xtabs(~Year+gcatP,data = lmbs))
```

```
##       gcatP
```

```
## Year   preferred+ preferred-
##   2013         16         82
##   2014         18        122
##   2015         15         52
##   2016         11         96
```

```r
chisq.test(lmb.LFP)
```

```
##
##  Pearson's Chi-squared test
##
## data:  lmb.LFP
## X-squared = 5.4469, df = 3, p-value = 0.1418
```

```r
(ps.P <- c(chisq.test(lmb.LFP[1:2,])$p.value,    ### 2013-2014
           chisq.test(lmb.LFP[c(1,3),])$p.value, ### 2013-2015
           chisq.test(lmb.LFP[c(1,4),])$p.value, ### 2013-2016
           chisq.test(lmb.LFP[2:3,])$p.value,    ### 2014-2015
           chisq.test(lmb.LFP[c(2,4),])$p.value, ### 2014-2016
           chisq.test(lmb.LFP[3:4,])$p.value))   ### 2015-2016

(p.val.P <- p.adjust(ps.P))
```

```
## [1] "13-14" "13-15" "13-16" "14-15" "14-16" "15-16"
```

```
##    Year p-value Adjusted p
## 1 13-14  0.5724      1.000
## 2 13-15  0.4378      1.000
## 3 13-16  0.2837      1.000
## 4 14-15  0.1212      0.606
## 5 14-16  0.6716      1.000
## 6 15-16  0.0498      0.299
```

The PSD-P of largemouth bass is not different for any years during 2013 - 2016 (Chi-Squared, $X^2 = 5.45$, p = 0.1418). The adjusted p-values show no difference in the PSD-P between years (2013 - 2016). The adjusted p-value between years 2015 - 2016 is the closest to being different (p = 0.3).